

# 1 Twitter in Gesellschaft und Forschung



**Abbildung 1:** Ursprünglicher, mittlerweile gelöschter Tweet von Justine Sacco auf Twitter. Bildquelle: Pilkington (2013).

„Going to Africa. Hope I don't get AIDS. Just kidding. I'm white!” (Sacco, 2013). Mit diesem Tweet verabschiedete sich Justine Sacco, Leiterin der Unternehmenskommunikation des New Yorker Medienunternehmens *InterActiveCorp*, am 20. Dezember 2013 vor ihrem Flug von London nach Kapstadt. Während sie sich wohl der Reichweite ihrer Äußerung nicht bewusst war, entstand binnen kürzester Zeit eine Welle der Empörung im Internet – ein sogenannter *Shitstorm*. Das Hashtag *#HasJustineLandedYet* war ein populärer Begriff auf Twitter (Ronson, 2015).

© Der/die Autor(en) 2016

F. Pfaffenberger, *Twitter als Basis wissenschaftlicher Studien*,

DOI 10.1007/978-3-658-14414-2\_1

Doch nicht nur auf Twitter wurde diese Meldung diskutiert – auch etablierte Printmedien thematisierten den Tweet in ihren Online-Ausgaben, wie die *New York Times* (Southall, 2013), der *Guardian* (Pilkington, 2013), *Die Welt* (Neumann, 2013) und der *Stern* (Noffke, 2013). Nachdem der öffentliche Druck auf Sacco stieg, erhielt sie noch am selben Tag die Kündigung. InterActiveCorp entschuldigte sich außerdem in einer Stellungnahme für das Verhalten seiner Mitarbeiterin (Stelter, 2013).

Dieser Fall zeigt, welche Bedeutung Twitter in der Gesellschaft einnehmen kann. Tweets sind mittlerweile ein beliebtes Mittel zum Verbreiten von Meldungen in Echtzeit: Beispielsweise kündigte der ehemalige griechische Finanzminister Yanis Varoufakis (2015) seinen Rücktritt auf Twitter an, ebenso wie *ProSieben* (2015) das Karriereende seines Entertainers Stefan Raab. Julia Klöckner informierte bei der Wahl zum Bundespräsidenten 2009 noch vor der offiziellen Bekanntmachung die Öffentlichkeit über den Wahlausgang (Boie, 2011). Der bisher weitverbreitetste Tweet wurde während der Oscar-Verleihung 2014 von der Moderatorin Ellen DeGeneres geschrieben und über 3,2 Millionen Mal<sup>1</sup> geteilt (DeGeneres, 2014). Er zeigte ein Gruppenfoto mehrerer Hollywood-Stars und wurde live – als Teil der Show – aufgenommen.

Twitter dient jedoch häufig nicht nur als Kanal zur einseitigen Bekanntgabe von Meldungen, sondern vor allem als Netzwerk zur *gegenseitigen* Information und Interaktion. Besonders im Umfeld der politischen Umwälzungen in Nordafrika und dem Nahen Osten spielte Twitter eine wichtige Rolle bei der Weitergabe von Informationen und der Koordination von Protesten (Bruns, Highfield, & Burgess, 2013; Lotan et al., 2011). Ägypten und die Türkei sperren beispielsweise gelegentlich einzelne Nutzer oder das ganze Portal – etwa bei politischen Ausschreitungen, kritischer Berichterstattung und neuerdings auch nach Anschlägen (Gadde, 2014; Kazim, 2015). Diese Maßnahmen sollen – aus der Sichtweise der jeweiligen politischen Regime – die Koordination von Protestbewegungen und die Weitergabe sensibler oder kritischer Informationen unterbinden.

Ähnlich reagierten Brüsseler Behörden: Während Razzien im Zuge der Pariser Terror-Anschläge im Dezember 2015 bat die Brüsseler Polizei die Twitter-User um eine selbst auferlegte Funkstille, damit potentielle Zielpersonen nicht gewarnt werden können (Police Fédérale, 2015). Viele Nutzer folgten der Bitte und veröffentlichten stattdessen Katzenbilder, um kompromittierende Tweets in der Masse untergehen zu lassen (Rogers, 2015). Bei den Terroranschlägen in Paris im November 2015 kamen viele Meldungen, Bilder und Videos zunächst nicht von der

---

<sup>1</sup> Stand: Februar 2016

Presse, sondern von Privatpersonen in sozialen Medien (Wendling, 2015). Diese Beispiele verdeutlichen die Präsenz und Bedeutung, die Twitter vor allem bei Katastrophen und Anschlägen, aber auch bei sportlichen und kulturellen Großereignissen haben kann.

Soziale Medien im Internet sind in den letzten Jahren rasant gewachsen. Die weite Verbreitung des Kurznachrichtendienstes Twitter mit seinen über 302 Millionen monatlich aktiven Nutzer/-innen und etwa 500 Millionen gesendeten Tweets pro Tag (Twitter, Inc., 2015k) steht nur beispielhaft für die Bedeutung moderner, internetbasierter Kommunikationskanäle. Statistiken zeigen, dass die knapp 56 Millionen deutschen Internetnutzer durchschnittlich 166 Minuten am Tag im Internet sind – etwa 39 Prozent nutzen dabei Online-Communities (ARD-Werbung Sales & Services GmbH, 2015). In diesem Zusammenhang sehen auch Politik, Medien und Unternehmen einen großen Nutzen in sozialen Medien: Sie eignen sich zur schnellen, einfachen, kostengünstigen und zeitunabhängigen Kommunikation mit den Wählern, Lesern oder Kunden.

Twitter bietet ein ausführliches, klar strukturiertes und frei zugängliches Datenset, welches sich sowohl für detaillierte, als auch für breit angelegte Datenanalysen gut eignet. So umfasst ein Datensatz rund 150 Metadaten – neben Tweet und Benutzerkonto auch Standortdaten, gewählte Sprache, Zeitzone oder die Vernetzung von Nutzer/-innen. In Abhängigkeit von der Tweetfrequenz ergibt sich ein ständig wachsender Datenpool. Aufgrund der standardisierten Struktur und der hohen Verfügbarkeit der Daten wird Twitter mittlerweile häufig als Datengrundlage für die wissenschaftliche Forschung in den unterschiedlichsten Disziplinen verwendet. Zwischen Januar 2007 und August 2015 finden sich auf *Scopus* allein 737 Forschungsarbeiten der Sozialwissenschaften, deren Titel den Begriff *Twitter* beinhalten, wobei die Anzahl an Artikeln in einem Jahr kontinuierlich ansteigt<sup>2</sup>. Daneben befassen sich auch andere Forschungszweige, wie die Medizin, Informatik, Ingenieurwissenschaften oder Psychologie mit Twitter (siehe auch Kapitel 2).

Je nach wissenschaftlicher Disziplin, Themensetzung, Datengrundlage, Rechenleistung und Budget ergeben sich dabei viele unterschiedliche Ansätze für die wissenschaftliche Verwendung von Twitter-Daten. Die Erhebung kann durch direktes Abgreifen von Tweets in Echtzeit, die Nutzung kostenfreier oder kostenpflichtiger Online-Aggregatoren oder durch den Erwerb vollständiger Datenbanken bei Datenhändlern vollzogen werden. Bei der Analyse besteht die Wahl zwischen der Nutzung (kostenpflichtiger) Online-Dienste und der Eigenauswertung mit Hilfe von Verarbeitungsprogrammen. Letztere können wiederum Lösungen

---

<sup>2</sup> Eigene Abfrage auf [scopus.com](http://scopus.com). Stand: August 2015. Suchterm: TITLE (twitter) AND (LIMIT-TO (SUBJAREA , "SOCI"))

*out of the box* oder Eigenentwicklung auf Basis freier Skriptsprachen sein, wie beispielsweise *Python*. Python ist eine frei verfügbare, umfangreiche und dennoch übersichtliche und leicht anwendbare Programmiersprache mit vielen Erweiterungsmöglichkeiten (s. Kapitel 3.2).

Trotz der Vielzahl an Studien und respektive an Ansätzen fehlt eine genauere Betrachtung und vor allem Bewertung typischer wissenschaftlicher Vorgehensweisen. Ziel dieser Arbeit ist deshalb eine Darstellung mehrerer methodischer Ansätze zur Erhebung, Speicherung und Analyse von Twitter-Daten auf Basis von Python. In der Arbeit sollen die verschiedenen Erhebungs-, Speicher- und Auswertungsverfahren näher betrachtet und anhand von Praxisbeispielen hinsichtlich ihrer Leistungsfähigkeit und ihres Nutzens für die Forschung bewertet werden. Darauf aufbauend wird die Eignung von Twitter als Quelle und Forschungsobjekt wissenschaftlicher Analysen erörtert, indem auch Einschränkungen und Herausforderungen aufgezeigt werden. Dabei liegt der Fokus nicht auf Vollständigkeit, sondern auf der Darstellung der Praktikabilität von Methoden anhand ausgewählter freier, kostenloser Programme.

Die Skizzierung einiger wissenschaftlicher Arbeiten unterschiedlicher Forschungsbereiche in Kapitel 2 soll nicht nur einen Überblick möglicher Verwendungszwecke für Twitter-Daten geben, sondern auch den Bedarf einer vergleichenden Darstellung möglicher Verfahren hervorheben. Kapitel 3 dient der Vermittlung von Grundlagen, indem Twitter, die spezielle Kommunikation auf diesem Portal und die Datenstruktur angesprochen werden. Zudem erfolgt eine kurze Einführung in die Programmiersprache Python. Kapitel 4 befasst sich schließlich mit der methodischen Abhandlung. Zuerst werden drei Ansätze zum Sammeln von Tweets vorgestellt: Abfragen über die beiden Programmschnittstellen<sup>3</sup> (APIs) *Streaming API* und die *REST APIs* sowie die Beschaffung von Twitter-Daten über Drittanbieter. Diese drei Möglichkeiten werden schließlich in Kapitel 4.1.4 gegenübergestellt und hinsichtlich ihrer unterschiedlichen Eignung für die Datensammlung bewertet.

Anschließend vergleicht Kapitel 4.2 zwei gegensätzliche Konzepte zum Speichern von Tweets: Das Speichern in Einzeldateien und das Einspeisen in Datenbanken, wobei ein Hauptaugenmerk auf dem Datenbanksystem *MongoDB* liegt. Daran anknüpfend erfolgt eine Betrachtung verschiedener Ansätze zum Auswerten der gesammelten Tweets (Kapitel 4.3). Zunächst thematisiert Kapitel 4.3.1

---

<sup>3</sup> Schnittstellen (Application Programming Interfaces, APIs) dienen beispielsweise der Kommunikation zwischen Datenbank und Nutzer. Über sie werden An- und Abfragen oder – allgemein gesagt – Befehle übermittelt und verwaltet. Die Kommunikation zwischen Schnittstelle und Endpunkt erfolgt dabei immer auf Quellcode-Ebene.

sinnvolle Vorverarbeitungsschritte zur Verbesserung der Datenqualität und Optimierung der Datenstruktur für eine automatisierte Analyse. Danach werden grundlegende, bereits in MongoDB integrierte Verfahren präsentiert und verglichen. Darüber hinaus gibt Kapitel 4.3.3 einen Einblick in die computergestützte Textanalyse, stellt grundlegende computerlinguistische Untersuchungen vor und schließt mit der Durchführung einer einfachen semantischen Analyse ab.

Schließlich beleuchtet Kapitel 5 die Eignung von Twitter als Quelle wissenschaftlicher Arbeiten, indem beispielsweise die Qualität und Zweckmäßigkeit der zur Verfügung gestellten Daten kritisch hinterfragt wird. Darüber hinaus sind auch rechtliche Bestimmungen zum Datenschutz und das Fehlen anerkannter Metriken von Bedeutung.

**Open Access** Dieses Kapitel wird unter der Creative Commons Namensnennung - Nicht kommerziell 4.0 International Lizenz (<http://creativecommons.org/licenses/by-nc/4.0/deed.de>) veröffentlicht, welche für nicht kommerzielle Zwecke die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Etwaige Abbildungen oder sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende oder der Quellreferenz nichts anderes ergibt. Sofern solches Drittmaterial nicht unter der genannten Creative Commons Lizenz steht, ist eine Vervielfältigung, Bearbeitung oder öffentliche Wiedergabe nur mit vorheriger Zustimmung des betreffenden Rechteinhabers oder auf der Grundlage einschlägiger gesetzlicher Erlaubnisvorschriften zulässig.

